Modeling Errors in Taxiing of Commercial Aircraft

Michael D. Byrne
Department of Psychology
Rice University
byrne@acm.org

Alex Kirlik
Aviation Human Factors
University of Illinois
kirlik@uiuc.edu

BACKGROUND

Taxiing from the runway to the gate or from the gate to the runway is the phase of commercial flight that is least automated. Since 1972, 11 runway accidents have claimed 719 lives, not counting the accident in Italy in the fall of 2001. Runway incursions have increased 15% per year for the last four years in the U.S. alone. The bulk of this increase is attributed to "pilot deviations," that is, errors made by the flight crew (captain and first officer). Thus, the problem of errors made during taxiing is a significant issue.

NASA Ames conducted several studies of taxiing in their high-fidelity flight simulator (e.g. Hooey, Foyle, Andre, & Parke, 2000). Flight crews flew into Chicago O'Hare airport with simulated dense fog with 700' visibility. The error rate was fairly high, with crews making at least one "major" error on 22% of the trials. These errors were classified on the basis of videotape protocols into three types:

• Communication errors. These are errors involving forming an incorrect intent or failures to understand or communicate route correctly. For example, if the route given to the crew was misheard or miscommunicated between the crewmembers, the error was classified as a communication error.

• Decision errors. These were errors involving making an incorrect decision at a turn point. For example, if the crew turned left where they were supposed to turn right, then the error was likely a decision error.

• Execution errors. These errors involved failing to correctly execute a turn maneuver or otherwise navigate an intersection. This included errors like following the wrong lead line or misinterpreting signage.

Follow-up studies showed that advanced cockpit technologies such as an electronic moving map (EMM) or a heads-up display (HUD) showing the correct route significantly reduced the error rate. However, these technologies are expensive and unlikely to make it into commercial cockpits anytime in the near future. The research here was motivated by the need to understand the ultimate sources of error and to try to generate predictive models, and possibly suggest ways of remediating error.

MODELING APPROACH AND MODEL

Our approach was to attempt to unify computational cognitive modeling in ACT-R/PM (e.g. Byrne & Anderson, 2001) with detailed environmental analysis (e.g. . We believe both of these methodologies are now advanced and formalized enough that they can be fruitfully merged. From the perspective of the ACT-R/PM model, the environmental analysis provides two things. First, it provides the model with a realistic environment of operation, such as realistic time constraints based on a model of physical aircraft dynamics. The environmental analysis and subject matter experts were also used to identify problem-solving and decision-making strategies, as well as setting parameters such as success rates.

The model is an ACT-R/PM model, based on the 4.0 version of ACT-R. This model is limited in scope in that we have only modeled the captain and not the first officer. Thus, the model has nothing to say about communication errors in this task. Also, many low-level details were omitted in the interest of time. In particular, the model does not include a detailed model of the motor actions necessary to steer the aircraft. The airport is modeled as a network of "rails" and the problem faced by the model is one of selecting the right "rail" for the plane to ride on.

The model has essentially two modes of operation: normal taxiing and intersection handling. During normal taxiing, i.e. following a single path between intersections, the model cycles through four critical maintenance tasks. These are:

• Look for incursions. The model scans the visual scene for anything that isn't supposed to be there.

• Maintain speed. Aircraft speed control is more complex than in other systems like automobiles. Constant throttle/brake will either lead to deceleration or acceleration, so speed must be monitored and throttle/

brake values adjusted on a regular basis. In the real task, this would also entail monitoring and correction of steering.

• Listen for holds. Planes are occasionally instructed by ground control to stop as they taxi to avoid collisions or alter traffic flow. The audio stream must thus be monitored to make sure a hold instruction is heard and followed.

• Update location. The model maintains a representation of its current location, and this location needs to be up-to-date as the plane moves in order for decision procedures to make appropriate turn decisions.

In the course of doing these tasks, the model may note that an intersection is approaching, which means the model must deal with the pending intersection. This decision is a two-stage process. The first step is deciding whether or not the intersection is one at which a turn is required. Second, if a turn is required, which turn should be taken?

Deciding on whether a turn should be made is fairly straightforward. For example, if the intersection is a "T", then a turn will clearly be required. Otherwise, the model generally relies on its memory for the taxi route to make this decision. As there was little evidence for this kind of error in the data, the model makes few errors here.

If a turn is required, then the model is forced to make a more complex decision, which is deciding which of the available turns to take. The model has several decision strategies available to it, and chooses between them on the basis of time remaining and strategy accuracy. Time is critical in this task, as the amount of time the model has to make the decision can be quite short, on the order of a few seconds. The amount of time available is dependent on a number of factors that are essentially external to the cognitive model, such as the braking dynamics of the aircraft, placement of taxiway signage, and taxiway geometry. For example, sharper turns require more slowing, and thus have a shorter decision horizon. Thus, having an accurate environmental model is critical in determining the cognitive model's behavior.

There are several decision-making strategies or heuristics available to the model. The fastest and least accurate is for the model to simply to attempt to retrieve from memory the direction of the turn. This is inaccurate for multiple reasons, the first of which is that the original taxi route is simply presented as a sequence of taxiway names which did not include turn directions. Second, of course, is that retrievals are themselves not perfectly reliable.

Another heuristic is essentially hill-climbing. That is, take the turn that seems to be most in the direction of the goal, which in this case is the gate. Our subject matter experts (SME) indicated that there are essentially two things they always try to know: current position and location of the gate. Thus, given these two pieces of information, this type of hill-climbing is always possible. This computation is not as rapid as a simple retrieval and is surprisingly accurate. We call this strategy the TT strategy for "toward terminal."

Another heuristic we derived from a combination of discussions with SMEs and an analysis of the task environment. If the airport is considered from a birds-eye view on an XY grid, progress down a taxiway changes the distance from the current location to the goal in either the X or Y direction (or both, but usually one is dominant). At the turn point, this heuristic selects the direction which will reduce the distance to the goal in the other axis. For example, if the plane is heading west and forced with a north-south choice, and the plane is currently north of the gate, the model will turn south. This is subtly different than simply turning toward the terminal, and is both more accurate and takes longer to compute. We term this the "XY" heuristic.

The "ultimate" decision strategy, and also the slowest, is for the model to consult the Jeppesen chart, which is a paper map of the airport. This is very slow, and while generally accurate, it is not certain to produce the correct result because of the difficulty in mapping from a 2D birds-eye map to a 3D self-centered world.

To gauge the accuracy of our intermediate heuristics, we had an SME indicate the typical taxi routes at a number of airports other than O'Hare, including Dallas, Miami, New York JFK, Atlanta, Los Angeles, Seattle, San Francisco, and Denver. From those routes, we determined the accuracy of the TT and XY heuristics at various airports. Across all airports, the TT strategy produces the correct decision 81% of the time and the XY heuristic 94% of the time. However, at O'Hare these heuristics are not as accurate, with TT producing the correct turn only 69% of the time. XY is still quite effective at O'Hare, producing correct responses 93% of the time. This reduction is due to the complex layout of O'Hare which entails a great deal of backtracking.

These effects are further magnified by the complexity of O'Hare, as the average number of turns that must be made across all of these airports is 4.1, while at O'Hare the average number of turns per route is 6.6. Interestingly, there are two turns in the corpus from the NASA experiment where both of these heuristics fail, and at both of these locations at least one error was made by the subjects.

The cognitive model is augmented with not only a model of the visual environment, which was based on the database used to drive the NASA simulator, but a model of the physical aircraft dynamics as well. This model was based on an approximate model of vehicle dynamics borrowed from Salvucci (2001) which is an automobile model. This model was adjusted based on thrust and weight specifications obtained from Boeing, then adjusted based on physics first principles given that the car approximation breaks down in certain places for an aircraft. This physical model is used to determining braking and acceleration response to inputs from the cognitive model, and was compared to empirical braking and acceleration data published by NASA (Cheng, Sharma, & Foyle, 2001).

ERROR BEHAVIOR

While the original experiment is not a sample large enough to produce data to which specific fitting would be appropriate, we do believe the model sheds some insight on the ultimate sources of error in this task. There are two major sources of error in this model. One is a fairly traditional source of errors in ACT-R models, which is failure of retrieval or mis-retrieval of various pieces of knowledge. This is not a particularly systematic error source, though it is sensitive to the amount of cognitive workload with higher workload yielding both slower and less accurate retrieval. A more systematic error source is the time/accuracy tradeoff in decision-making. Strategies that take less time are also less accurate, and the task often has entails tight time constraints. There are other secondary sources of error, such as the model failing to see a necessary sign in the time provided, which is really a secondary effect of the time pressure.

The model thus provides at least an explanation of the decision errors described in the original NASA data set. We believe that some of the execution errors could be modeled in a straightforward manner by degrading the perceptual inputs to the model. While we have not yet done that, there is not an in-principle barrier to doing so, as ACT-R/PM contains mechanisms for handling degraded perceptual input.

FUTURE WORK

Pieces of this work are still unfinished. First of all, we would like to get the system optimized enough to run Monte Carlo simulations to determine parameter sensitivity. As mentioned, we would also like to experiment with degraded perceptual inputs. Further examinations that might be possible are explorations of other decision-making strategies and adding a second model, which would be a model of the first officer in this task.

ACKNOWLEDGEMENTS

We would like to thank Brian Webster, Michael Fleetwood, and Christopher Fick for their contributions to this project. This research was supported by NASA cooperative agreement NCC-1219. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

REFERENCES

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an “aid” can (and should) go unused. *Human Factors, 34(2),* 221-242.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review, 108,* 847–869.

Cheng, V. H. L., Sharma, V., and Foyle, D. C. (2001). Study of aircraft taxi performance for enhancing air-port surface traffic control. Manuscript accepted for IEEE-ITS publication.

Hooey, B. L., Foyle, D. C., Andre, A. D., and Parke, B. (2000). Integrating datalink and cockpit display technologies into current and future taxi operations. *Proceedings of the AIAA/ IEEE/SAE 19th Digital Avionics System Conference.*

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies, 55,* 85-107.